Emotion Sensing for Mobile Computing

Jiayu Shu, Mangtik Chiu, and Pan Hui

The authors introduce how mobile and wearable devices work as "emotion sensors" by leveraging their sensing, computing, and communication capabilities, which help monitor people's mental health, facilitate social interactions, and improve user experience.

Abstract

Emotion is a complex mental state that guides people's thoughts and behaviors. The ability to recognize one's own emotions and to understand others' emotions helps people manage their personal lives and social relations more successfully. Meanwhile, mobile and wearable devices have become indispensable tools for most of the population. They provide great opportunities for building emotional intelligent applications. In this article, we introduce how mobile and wearable devices work as "emotion sensors" by leveraging their sensing, computing, and communication capabilities, which help monitor people's mental health, facilitate social interactions, and improve user experience. We first describe a general emotion sensing framework that consists of sensing, inferring, and responding steps. After reviewing the state-of-the-art systems on this topic, we summarize and discuss various sensing modalities enabled by mobile and wearable devices. We then describe the widely used inferring procedures and methods. Additionally, we present three solutions of facial expression recognition on smartphones as a case of emotion inference. Finally, we introduce emerging emotion sensing applications, and discuss challenges and opportunities.

Introduction

Emotion is a complex state of feeling that results in physical and psychological changes. It is usually associated with one's personality, mood, circumstances, and relationships with others. As the decisions we make and the actions we take are all influenced by the emotion we are experiencing, it is important to perceive and regulate our emotions. The ability to interpret and react to others' emotions is also essential, because it helps us respond appropriately and communicate effectively in various social situations.

Since emotion plays a critical role in people's daily lives and social communication, researchers have been developing machines and systems that are able to recognize, respond to, and even express emotion. Similar to the way in which humans perceive others' emotions, machines can recognize human emotions via people's behavioral responses. Therefore, facial expressions and speech have been widely analyzed as they contain rich patterns that reflect people's emotions [1]. While they are obvious cues of emotion, the tasks involved in recognizing emotions are power-hungry and computationally intensive. On the other hand, as people feel emotion and experience physiological reactions at the same time, machines can also recognize human emotions by analyzing physiological signals. However, the acquisition of such signals usually requires sensors that are either cumbersome or invasive when attached to individuals.

In recent years, the number of mobile and wearable device users has grown rapidly. From smartphones, smart watches, to smart wristbands and smart glasses, these devices continuously develop in capacity, performance, and intelligence, which offer new possibilities and opportunities for emotion sensing. First, different types of emotion-related data can be collected simultaneously in an unobtrusive way. Mobile and wearable devices are embedded with a number of physical sensors that are able to detect a user's motion, monitor physiological signals, capture images and videos, and record surrounding sounds [2]. Moreover, there are virtual sensors that can acquire a user's usage statistics, since the device constantly stays within the user's proximity and interacts with the user at an unprecedented frequency. The physical and virtual sensors can work together to collect emotion-related data without making the user feel uncomfortable or be invaded. Second, data analysis can be performed flexibly. Higher computation capabilities of mobile and wearable devices now allow them to perform most data analysis tasks locally to get results promptly. Alternatively, the task can be offloaded to neighboring devices, to edge elements, or to the cloud, where collaborative data analysis approaches can be used [3]. Finally, emotional intelligence of mobile and wearable devices is important for a variety of applications and scenarios. For example, it is challenging to measure and quantify quality of experience (QoE). An emotion sensing function could be a feasible solution for developers and service providers to infer QoE based on users' emotional states sensed from their devices, and then take action to improve user experience and service quality.

In this article, we aim to give readers an overview of the trends in mobile emotion sensing, especially the state-of-the-art emotion sensing methods and emerging applications on mobile and wearable devices. To this end, we first describe an emotion sensing framework for mobile and wearable devices that consists of three steps: sensing, inferring, and responding. Then we discuss sensing modalities, including individual status, device usage, application content, and textual information. Next, we describe the inferring procedures, namely data pre-processing,

Digital Object Identifier: 10.1109/MCOM.001.1800834

Jiayu Shu, Mangtik Chiu, and Pan Hui are with Hong Kong University of Science and Technology; Pan Hui is also with the University of Helsinki.

0163-6804/19/$25.00 © 2019 IEEE

feature extraction, and model training. To further show the feasibility and characteristics of different inferring methods, we present three solutions of facial expression recognition on smartphones. Finally, we introduce emerging emotion sensing applications, and discuss challenges and opportunities in this area.

Emotion Sensing Overview

Emotion sensing mainly consists of three stages as shown in Fig. 1: sensing, inferring, and responding.

Sensing

The sensing stage collects emotion-related data. Human emotions are expressed and can be observed in many ways, including behavioral responses (e.g., smiling, avoiding social contacts) and physiological reactions (e.g., racing heartbeat, rapid breathing). As mobile devices are equipped with a number of physical sensors such as cameras, microphones, accelerometers, touch panels, GPS sensors, and proximity sensors to see and feel the external world, they are suitable platforms for collecting emotional responses without interrupting users. Moreover, specially designed wearable devices like wristbands and smart watches with sensors such as galvanic skin response (GSR) and electromyography (EMG) sensors are able to detect certain physiological signals in an unobtrusive way. In addition, people's daily interactions with devices also contain rich information, such as which applications people use, with whom people communicate, and so on.

Inferring

The inferring phase aims to recognize emotions by analyzing collected data. To describe and measure human emotional states, psychologists have proposed several emotion models, which can be classified into discrete and dimensional models. In discrete models, happiness, sadness, disgust, fear, surprise, and anger are a set of basic emotions that are usually detected, as they are believed to be universally experienced across cultures. The advantages of discrete models include their simplicity and conformity with people's description of emotions in daily life. However, not all emotions can be described with a single word due to the fuzzy boundaries between emotional states. On the other hand, dimensional emotion models describe emotions in multiple dimensions. The most prominent one used in emotion sensing systems is the Circumplex model, which measures emotion in a two-dimensional circular space spanned by valence and arousal. Arousal represents activation levels (high vs. low), and valence stands for pleasantness (positive vs. negative). As dimensional models spread across multiple axes, they are able to comprehensively represent and distinguish between many more emotional states compared to discrete models. However, the complexity of dimensional models brings difficulties in the emotion labeling and inference process.

Usually, machine learning methods are used to estimate the most likely emotional state. In machine learning, such a learning paradigm is called supervised learning, which is to map a new input to an output based on a set of training input-output pairs. The task can be performed locally on a device, in the cloud, or on other platforms where computing resources are available.

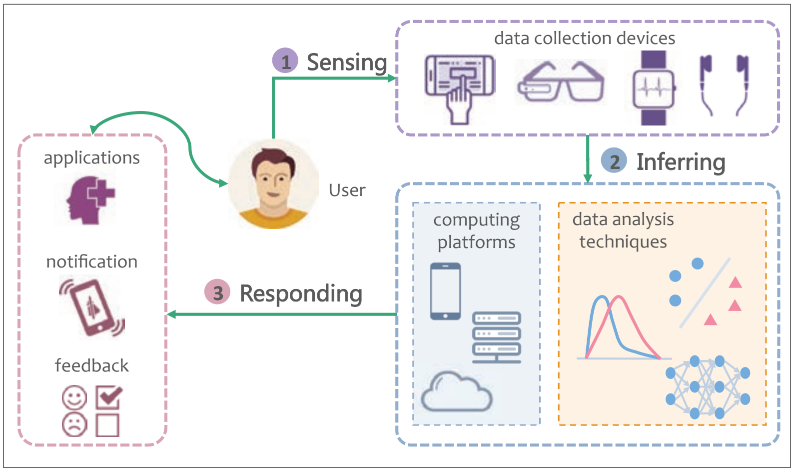

Figure 1. Emotion sensing framework for mobile and wearable devices. In the sensing stage, emotion-related data are collected from various sensors. In the inferring stage, data are analyzed to infer emotional states. In the responding stage, applications respond to detected emotions.

Responding

The final stage, responding, is concerned with utilizing the emotion inference results. The emotional states of a specific user provide useful input for various applications. For example, detecting users' emotional states while they are playing games or listening to music will help applications offer more personalized experiences. By keeping track of users' emotional states, devices can monitor their mental health, notifying them if there are any potential mental problems. Meanwhile, presenting the emotion inference results to users in an appropriate way is also an important issue. For instance, there are applications that help people with visual impairments to engage in conversations by informing them of others' emotions. How to present results to users, by generating either an audible signal or vibration, requires careful consideration. Moreover, users can provide feedback to the emotion sensing module to further improve the inference performance, especially if the inferred results are users' real emotional states.

Sensing: Collecting Emotion-Related Data

In this section, we discuss four major sensing modalities as summarized in Table 1, namely individual status, device usage, application content, and contextual information.

Individual Status

Individual status refers to behavioral and physiological responses that can be detected by physical sensors on devices, including facial expressions, speech, physiological signals, body gestures, and activities.

Facial Expression and Speech: Facial expression is a gesture executed with the facial muscles that conveys an individual's emotions. Both spatial information on one single image (e.g., wide open

Sensing modality	Data	Description	Properties
Individual status	Facial expression and speech	Gesture executed with facial muscles and vocalized form of communication	Data has consistent and repeated patterns. The inferring process is power-hungry and computationally intensive.
	Physiological signal	Electrical and non-electrical signals originating from the central and peripheral nervous system	Data collection requires dedicated or cumbersome sensors. Data are more spontaneous and less controllable.
	Body movement and activity	Motion and action of the body	Data collection is unobtrusive. Data is used with other modalities to infer emotion.
Device usage	Social interactions and routine activities	Social activities (e.g., calls, messages) and daily interactions (e.g., app usage)	Data collection is unobtrusive. Data is individual-dependent.
	Touch and typing	Screen touch behaviors and typing characteristics	Data collection is unobtrusive. Data is individual-dependent and application-dependent.
Application content	Image, text, emoticon	Visual, audio, and textual content of apps	Data collection is unobtrusive.
Contextual information	Time, location, weather, proximity	Circumstances that form the setting for events or situations	Data collection is unobtrusive. Data is used with other modalities to infer emotion.

Table 1. Description and properties of four major emotion sensing modalities.

mouth) and temporal information (e.g., quickly blinking eyes) from a sequence of images can be transformed to facial features for emotion inference. Speech is a vocalized form of communication based on the syntax combination of different vocabularies. Speed, tone, stress, and volume all provide cues for recognizing individuals' emotional states. Among various emotion-related data, facial expressions and speech have been widely studied for years, mainly because they reveal one's emotions directly with consistent and repeatable patterns [1]. Although they can easily be captured by devices with cameras and microphones anywhere and anytime, the power consumption and computational cost of collecting and analyzing visual and audio data still pose a burden on mobile devices.

Physiological Signal: Physiological signal refers to any signal that can be generated, measured, and monitored in living organisms. Common physiological signals of human beings that have been used for emotion recognition include electrical signals such as electroencephalography (EEG), electrocardiography (ECG), electromyography (EMG), galvanic skin response (GSR), as well as non-electrical ones such as respiration rate, blood pressure, and heart rate. As with other time series data, features from both the time and frequency domains are extracted to infer emotion. Although most physiological signals need to be detected using sensors, which are obtrusive and cumbersome, the data is more spontaneous and less controllable compared to visual and audio expressions, since they originate from the central or peripheral nervous system that cannot be triggered consciously or intentionally [4]. Now the trend is to develop more compact sensors and advanced data acquisition techniques to capture physiological signals in an unobtrusive way. For example, photoplethysmography is a technology used in smart watches to measure heart rate, which illuminates the skin and measures the amount of light that is scattered by blood flow.

Body Movement and Activity: Body movements and activities are people's everyday movements and actions, such as walking, running, cycling, and sleeping. As studies indicate that relations exist between people's physical activity levels and their mental states [5, 6], motion features like acceleration, step duration, speed, and orientation are extracted from raw sensor data (e.g., accelerometer, gyroscope, GPS) to help infer emotion. It provides body cues that help better discriminate between the expressed emotion, and at the same time does not disrupt the normal use of the device.

Device Usage

People use mobile and wearable devices at an unprecedented rate nowadays. Their interactions with devices include social activities, such as making phone calls and sending messages, and daily routine activities, such as browsing the Internet and using different categories of applications. Studies have demonstrated the relationship between personal emotions and usage patterns of smartphones [7, 8]. Furthermore, researchers have found that users' interactions with devices such as typing characteristics [9] and touch screen behaviors [10] may reveal their current emotional states as well.

The advantage of analyzing usage patterns and device interactions is that the data can be obtained easily without interrupting the user. However, analysis of this type of data is more complicated and is still in the exploratory stages, as people have different usage habits. Besides, touch and typing behaviors vary among applications. Therefore, inferring users' emotions based on usage patterns and device interactions is individual-dependent and application-dependent, which requires long-term data collection.

Application Content

Applications installed on smartphones like social networking ones provide a convenient and timely channel for users to communicate with others and express their thoughts and feelings. Users generate visual, audio, and textual content when using these applications. For example, when a person is happy or sad, he/she may post an image on Instagram, update his/her status on Facebook, or send an emoji to his/her friends. Such information directly implies the emotions of the user. As a result, techniques such as image emotion recognition and text sentiment analysis can be applied to infer a user's emotions by leveraging the application content from the user.

Contextual Information

Contextual information helps describe circumstances that form the setting for events or situations. Researchers have been using context as extra information to help infer emotions, as emotions are affected by specific context [7, 8]. Contextual information (e.g., a user's current location) can be obtained by GPS. The crowd density in a user's proximity can be detected by Bluetooth. Time and weather are two other contextual elements that can be considered. Finally, the social context may also help, as in some situations, a group of people would be in similar emotional states. In short, although context is not a direct indicator of emotion, it provides supplementary information.

Inferring: Analyzing Emotion-Related Data

In this section, we discuss data analysis methods that are used to infer emotions. As a typical classification problem, there are several major steps: pre-processing, feature extraction, and model training.

Data Pre-Processing

Data pre-processing is a preparation step before extracting any features or feeding the data into any model. The goal is to get data of high quality out of the raw data that are generally incomplete, noisy, inconsistent, and redundant, which finally avoids misleading results. The specific data pre-processing techniques to be applied are largely dependent on the type of data. For instance, for facial images, histogram equalization reduces the influence of illumination, and face alignment eliminates the effect of shooting angle. For physiological signals, band-pass filters are used to remove the noise.

Feature Extraction and Model Training

The objective of feature extraction is to find distinctive attributes from the initial set of data that can best represent the data to distinguish between different emotional states. Researchers usually use domain knowledge to create features, the process of which is called feature engineering. Some features have numerical presentations, such as the distance between two points in an image, the frequency of one event during a certain period of time, or the average value of the signal. Features can also be in categorical form, such as weather conditions. In some cases, however, not all manually extracted ones are relevant to emotions. Hence, feature reduction methods are applied to get a minimum possible subset of attributes, while the performance of the classification model will not be compromised. Then a classifier is trained with feature vectors and the corresponding emotion classes as input-output pairs. Popular basic classifiers used in emotion sensing include logistic regression, decision tree, support vector machines (SVMs), Bayesian network, and random forest. These models try to solve the classification problem from multiple perspectives and different principles; therefore, they differ in complexity, accuracy, and processing time with the same training data. To select a suitable model that achieves the desired results with the available resources, one could try and select the model with the best performance.

On the other hand, deep learning is known for learning features automatically. The multi-layered architecture learns multiple levels of representations. For example, convolutional neural networks (CNNs) outperform the aforementioned algorithms in classification accuracy in most computer vision tasks, such as facial expression recognition.

Facial Expression Recognition on a Smartphone as a Case

When deploying emotion sensing applications on mobile and wearable devices, multiple factors should be taken into consideration, including the overall system performance, computational cost, response time, user experience, and so on. Here, as a case, we implemented facial expression recognition systems on smartphones to infer one of the six basic emotions when the camera captures a face, using both the basic machine learning and deep learning methods.

Following the way a basic classifier is trained, we extract features manually, which are the Euclidean distances for each of the 51 facial landmarks between a neutral face and a face expressing emotion, as shown in Fig. 2a. We then train an SVM model with the CK+ dataset, which contains 593 grayscale videos of 123 subjects.

As for the deep learning method, we use the popular VGG-16 model, changing the convolution kernel size and number of layers to adapt to the task (Fig. 2b). We train the model using 27,000 images of the fer2013 dataset for 10,000 iterations. We also use the emotion recognition application programming interface (API) from Microsoft Azure Cognitive Services (MS-API). Once a facial image is uploaded, the API will return the classification result based on a deep CNN.

Table 2 lists the performance of three systems running on a OnePlus 5 smartphone, with a Snapdragon 835 CPU and 8 GB RAM. The validation accuracies are obtained using the images from their validation sets (i.e., images in the training set but not used in the training phase), while the evaluation accuracies are all obtained from the TFEID dataset. The results indicate that:

- With sophisticated pre-processing steps and features, basic machine learning methods may achieve comparable accuracies with deep learning methods under certain conditions.
- Deep learning methods simplify the inference process, but the accuracy largely depends on the quantity and quality of the training data.
- The cloud service is a good choice for applications that require high accuracy and less computational burden, but can tolerate response delay. With the incoming fifth generation (5G) technologies, the response time can be further decreased.

Data pre-processing is a preparation step before extracting any features or feeding the data into any model. The goal is to get data of high quality out of the raw data that are generally incomplete, noisy, inconsistent, and redundant, which finally avoids misleading results. The specific data pre-processing techniques to be applied are largely dependent on the type of data.

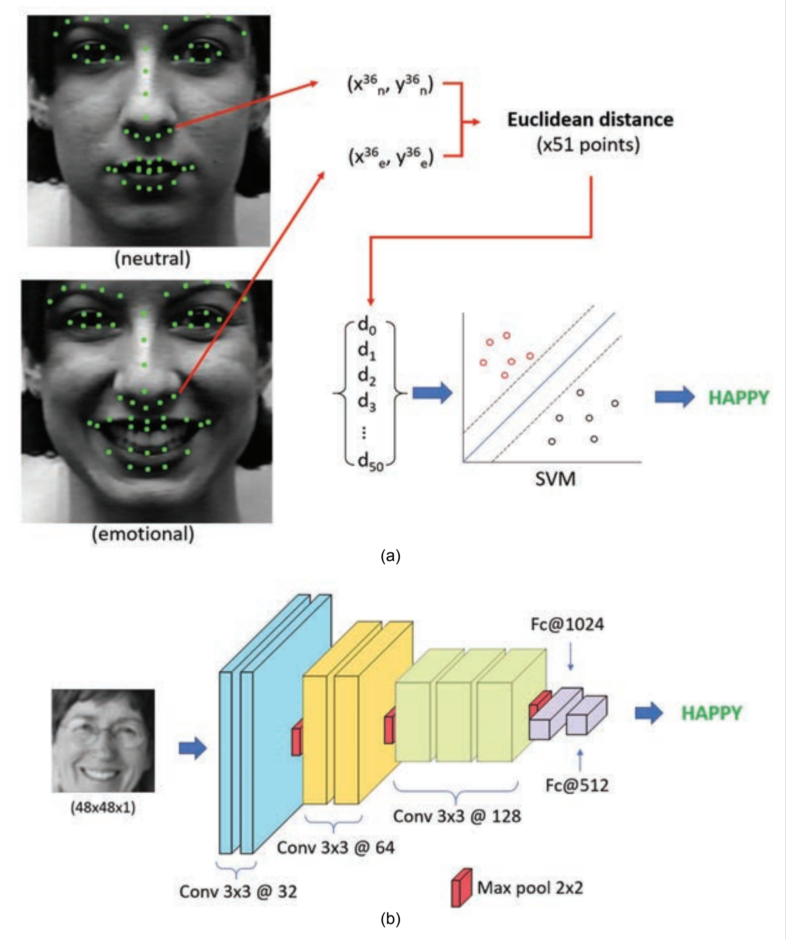

Figure 2. a) Handcrafted features are first extracted from neutral and emotional faces, pipeline of emotion classification with SVM; b) CNN automatically learns features for the classification task, pipeline of emotion classification with CNN.

Multimodal Fusion

As people's emotional responses are multi-dimensional, we can leverage multiple types acquired simultaneously on mobile and wearable devices. The process of integrating data and knowledge from several sources to produce more accurate and consistent results is called multimodal fusion. Current fusion strategies for emotion recognition can be classified into feature-level, model-level, and decision-level fusion [11]. Feature-level fusion simply concatenates features extracted from single-modal data to construct a joint feature vector. For instance, physiological features and device usage features can be combined together before the model training. On the other hand, model-level fusion combines knowledge from different views by learning a shared feature across multimodal data. Compared to feature-level fusion, it is able to capture the correlations and interactions among multimodal data. Decision-level fusion ensembles the inference results from single modalities to make a final decision. It has advantages when data from different modalities have different formats and can be modeled independently, such as facial expressions and physiological signals.

Applications

Emotion sensing has multiple applications on mobile and wearable devices. In this section, we introduce four major aspects, ranging from recognizing a user's emotions to understanding others' emotions.

Mental Health Monitoring

Measuring and monitoring individual mental states for the purpose of prevention, notification, and evaluation are significant but challenging. Traditionally, psychiatrists and psychologists analyze people's emotional states and mental health by means of observation, self-reporting, and questionnaires. However, these methods may not be able to capture unconscious emotional responses as people are aware of being observed, and might lead to biased results. Therefore, mobile and wearable devices help to measure and monitor emotional responses in an unobtrusive way, improving the reliability of measurements. For example, EmotionSense is a mobile sensing system that automatically detects users' emotions from speech by running classifiers locally on smartphones for social psychology studies [12]. Bogomolov *et al.* recognized stress based on device usage (social activities), proximity information, weather, and personality traits [7]. Wang *et al.* monitored and assessed the mental health changes for patients with schizophrenia using smartphone sensor data, including activities, device usage, and context [6]. Shuai *et al.* proposed a framework to identify potential cases of social network mental disorders using online social behaviors [13].

User Experience Improvement

Measuring QoE, such as frustration or engagement level when people are using applications on their mobile devices, helps networking service providers and application developers to improve service quality and user experience. For instance, the detection of frustration induced by system response delay can inform developers about the influence of networking conditions and choice of algorithms. Thus, Taylor *et al.* developed a model to predict multiple levels of user frustration from physiological responses using an armband, a heart rate detector, and a finger clip skin conductivity sensor when users are playing an interactive game [14]. Huynh *et al.* measured the engagement level of a smartphone game player using a smart wristband to measure touch behaviors, physiological signals, and a camera to capture upper-body skeletal motion data [15].

Mobile Intelligent Assistance

Intelligent assistance refers to the use of intelligent agents that help individuals perform tasks or services. In recent years, intelligent agents have evolved from those that respond to keyboard and mouse input into voice and video input, with the ability to recognize and react to human emotions. Examples include voice assistants on smartphones that can understand the current emotional state of the user and respond appropriately, and intelligent in-vehicle information systems that help drivers to avoid dangerous driving situations by detecting drivers' stress levels in real time.

Social Interaction

Interacting with others is a significant part of people's social lives. However, it is difficult for people with autism to respond to some non-verbal forms

	Validation Accuracy (%)	Evaluation Accuracy (%)	Inferring Time (ms)	CPU Usage (%)	RAM Consumption (MB)	Running Conditions
SVM	75.3	58.7	770.0	21.4	245.2	Requiring a pair of facial images and pre-processing
CNN	60.0	50.2	52.8	25.2	242.8	Running end-to-end locally on device
MS-API	N/A	70.4	914.8	21.8	149.2	Running end-to-end, requiring Internet connection

Table 2. Performances of three facial expression recognition solutions on an Android smartphone with specific running conditions.

of communication, such as facial expressions, gestures, and eye contact. Thus, many of them are unable to understand or interpret others' emotions. In addition, people with visual impairments have difficulty in getting emotional cues from others' facial expressions when engaging in conversations. Therefore, real-time emotion recognition systems running on devices like smart glasses can help people detect others' emotional changes to promote interpersonal communication by analyzing others' facial expression and speech. Also, people's emotional changes may not be observed sometimes. Accurate online emotion sensing allows people to interact offline with their friends when they actually need emotional support.

Challenges and Opportunities

In this section, we discuss several challenges and opportunities of emotion sensing.

First, people do not experience a pure form of a single emotion most of the time. A person may feel excited and nervous at the same time, for example, when a new task is assigned. Therefore, describing and inferring complex emotional states that contain multiple basic emotions at one time is essential for a more accurate and comprehensive emotion inference service.

Moreover, people experience and express emotion quite subjectively. The intensity of people's outwardly expressed emotions, such as facial expressions, differ from each other. The difference is more apparent for behavioral responses like device usage, which suggests a more personalized inference model that best interprets the emotional changes of a specific person. However, collecting sufficient personal data with corresponding labels is time-consuming. As a result, we need techniques that can transfer knowledge from various large-scale collective data to a personalized inference model that can infer a user's emotion in a cold-start condition.

In addition, privacy issues challenge the deployment of emotion sensing applications on personal mobile and wearable devices. Applications need to collect data from their users for training models. Devices may offload the inferring task to the cloud or other devices due to their limited computational capability or battery power. As a result, users' data are at risk of being disclosed. Location, device usage, activity data, calls, and messages all contain information that may reveal a user's identity. Therefore, a uniform privacy protection standard is required to control and monitor the use of personal data by emotion sensing functionalities.

Conclusion

Nowadays, sensors on mobile and wearable devices extend human senses. Meanwhile, the development of cloud and edge computing allows more flexible data analysis. By combining the sensing, computation, and communication capabilities of mobile and wearable devices, they can help perceive human emotions for various mobile computing and communication applications, such as monitoring mental health and measuring QoE. In this article, we provide an overview of the methods and trends of emotion sensing on mobile and wearable devices. Specifically, we introduce the state-of-the-art emotion sensing modalities, inferring methods, and emerging applications. We also discuss challenges and opportunities of mobile emotion sensing, including inference of complex emotions, a personalized inference model, and privacy issues. Although emotion sensing technologies are not yet mature, we believe multiple emotion-related data collection channels, advanced emotion inference methods, and various application scenarios will promote the development of emotion sensing in the near future.

Acknowledgment

This research has been supported in part by project 16214817 and G-HKUST604/16 from the Research Grants Council of Hong Kong and the 5GEAR project of the Academy of Finland.

References

[1] Z. Zeng *et al.*, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, 2009, pp. 9–58.

[2] N. D. Lane *et al.*, "A Survey of Mobile Phone Sensing," *IEEE Commun. Mag.*, vol. 48, no. 9, Sept. 2010.

[3] M. Chen *et al.*, "EMC: Emotion-Aware Mobile Cloud Computing in 5G," *IEEE Network*, vol. 29, no. 2, Mar./Apr. 2015, pp. 32–38.

[4] S. Jerritta *et al.*, "Physiological Signals Based Human Emotion Recognition: A Review," *Proc. IEEE 7th Int'l. Colloquiem Signal Processing and Its Applications*, 2011, pp. 410–15.

[5] F. Gravenhorst *et al.*, "Mobile Phones as Medical Devices in Mental Disorder Treatment: An Overview," *Personal and Ubiquitous Computing*, vol. 19, no. 2, 2015, pp. 335–53.

[6] R. Wang *et al.*, Crosscheck: Toward Passive Sensing and Detection of Mental Health Changes in People with Schizophrenia," *Proc. 2016 ACM Int'l. Joint Conf. Pervasive and Ubiquitous Computing*, 2016, pp. 886–97.

[7] A. Bogomolov *et al.*, "Daily Stress Recognition from Mobile Phone Data, Weather Conditions and Individual Traits," *Proc. 22nd ACM Int'l. Conf. Multimedia*, 2014, pp. 477–86.

[8] X. Zhang *et al.*, "Moodexplorer: Towards Compound Emotion Detection via Smartphone Sensing," *Proc. ACM Conf. Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, 2018, p. 176.

[9] B. Cao *et al.*, "Deepmood: Modeling Mobile Phone Typing Dynamics for Mood Detection, " *Proc. 23rd ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*, 2017, pp. 747–55.

[10] Y. Gao, N. Bianchi-Berthouze, and H. Meng, "What Does Touch Tell Us about Emotions in Touchscreen-Based Gameplay?" *ACM Trans. Computer-Human Interaction)*, vol. 19, no. 4, 2012, p. 31.

[11] S. Poria *et al.*, "A Review of Affective Computing: From Unimodal Analysis to Multimodal Fusion," *Info. Fusion*, vol. 37, 2017, pp. 98–125.

[12] K. K. Rachuri *et al.*, "Emotionsense: A Mobile Phones Based Adaptive Platform for Experimental Social Psychology Research," *Proc. 12th ACM Int'l. Conf. Ubiquitous Computing*, 2010, pp. 281–90.
[13] H.-H. Shuai *et al.*, "A Comprehensive Study on Social Network Mental Disorders Detection via Online Social Media Mining," *IEEE Trans. Knowledge and Data Engineering*, vol. 30, no. 7, 2018, pp. 1212–25.
[14] B. Taylor *et al.*, "Using Physiological Sensors to Detect Levels of User Frustration Induced by System Delays," *Proc. 2015 ACM Int'l. Joint Conf. Pervasive and Ubiquitous Computing*, 2015, pp. 517–28.
[15] S. Huynh *et al.*, "Engagemon: Multi-Modal Engagement Sensing for Mobile Games," *Proc. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, 2018, pp. 13.

Biographies

Jiayu Shu (jshuaa@connect.ust.hk) received her Ph.D. degree from the Department of Computer Science and Engineering, Hong Kong University of Science and Technology in 2019, and her B.Eng. degree from the School of Electronic Science and Engineering, Nanjing University in 2014. Her research interest lies at the intersection of mobile computing and computer vision. She designs and builds intelligent mobile systems to address real-life problems and assist humans in daily activities. Her research topics include visual privacy protection, augmented reality, and emotion sensing.

Mangtik Chiu (mtchiu@connect.ust.hk) is pursuing his Master's degree in computer science at the University of Illinois, Urbana-Champaign. He received his Bachelor's degree in computer science and engineering at Hong Kong University of Science and Technology. His research interests include computer vision and deep learning.

Pan Hui [F] (panhui@cse.ust.hk) received his Ph.D. degree from the Computer Laboratory, University of Cambridge, and his M.Phil. and B.Eng. degrees from the Department of Electrical and Electronic Engineering, University of Hong Kong. He is the Nokia Chair in Data Science at the University of Helsinki and director of the HKUST-DT Systems and Media Lab at Hong Kong University of Science and Technology. He also served as a distinguished scientist at Telekom Innovation Laboratories (T-labs), Germany, and an adjunct professor of social computing and networking at Aalto University, Finland. He is an ACM Distinguished Scientist and a member of the Academia Europaea.